

UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR



**Grado en Ingeniería de Tecnologías y Servicios de
Telecomunicación**

TRABAJO FIN DE GRADO

**DETECCIÓN DE VEHÍCULOS MEDIANTE ANÁLISIS DE
IMÁGENES**

Juan Ignacio Bravo Pérez-Villar
Tutor: Luis Salgado Álvarez de Sotomayor
Ponente: Jesús Bescós Cano

JULIO 2015

Resumen

El objetivo de este Trabajo de Fin de Grado es la implementación y análisis de un sistema capaz de segmentar las imágenes provenientes de la cámara situada en la parte frontal de un vehículo para la detección de vehículos, pavimento y líneas de carretera. El sistema usa información *a priori* de la escena para generar modelos estadísticos de las regiones de interés, que serán optimizados a través del algoritmo *Expectation-Maximization* para clasificar las imágenes de la secuencia de vídeo bajo el marco de la teoría de decisión bayesiana.

Se han estudiado e implementado las diferentes etapas del sistema, donde se han propuesto mejoras a algunas de ellas. Posteriormente, para evaluar el rendimiento del algoritmo y la validez de las mejoras se ha diseñado un conjunto de pruebas específico para cada clase, donde se comparan tres versiones diferentes del sistema con un conjunto de imágenes de referencia etiquetadas a mano para este propósito. Finalmente se discuten razonadamente los resultados obtenidos y se marcan las pautas para el trabajo futuro.

Palabras clave:

Detección de vehículos, líneas de carretera, transformación inversa de perspectiva, clasificación bayesiana, algoritmo *Expectation-Maximization*, segmentación.

Abstract

The goal of this Final Degree Thesis is the implementation and analysis of a system able to segment a video flow generated by a forward looking camera placed in the frontal of a vehicle, with the aim of detecting vehicles, pavement and lane marks. The system handles prior information of the scene to generate statistical models of the regions of interest, which will be optimized through the Expectation-Maximization algorithm to classify the images using the Bayesian decision theory.

All the stages of the system have been studied and implemented, contributing also with proposals to improve the performance of some of them. Subsequently, to evaluate the performance of the algorithm and the validity of the improvements, we have designed a group of specific tests for each class, where three different versions of the systems are compared against a ground-truth created specifically for this purpose. Finally all results are discussed and some guidelines for future work are given.

Keywords:

Vehicle detection, lane marks, inverse perspective mapping, Bayesian classification, Expectation-Maximization algorithm, segmentation.

Agradecimientos

En primer lugar agradecer a mi tutor Luis Salgado la oportunidad de realizar este trabajo con él y enseñarme que siempre se puede dar algo más. Gracias a esas personas con las que he compartido estos cuatro años y han hecho más llevadera esta aventura.

Finalmente agradecer a mi familia todo el apoyo dado, en especial a mi madre quien siempre ha sabido hacerme ver el lado bueno de las cosas.

Juan Ignacio Bravo Pérez-Villar

Julio, 2015

A mi padre,

Índice general

Resumen.....	I
Palabras clave:.....	I
Abstract.....	III
Keywords:.....	III
Agradecimientos	V
Índice general.....	IX
Índice de tablas.....	XI
Capítulo 1: Introducción	1
1.1 Motivación.....	1
1.2 Objetivos.....	2
1.3 Organización de la memoria	2
Capítulo 2: Estado del Arte	3
2.1 Introducción	3
2.2 Transformada de Perspectiva Inversa	4
2.3 Detección de líneas de carril	5
2.4 Detección de vehículos.....	7
Capítulo 3: Arquitectura del sistema.....	9
3.1 Introducción	9
3.2 Decisión bayesiana	11
3.3 Funciones de verosimilitud	12
3.4 Estimación de parámetros.....	14
3.4.1 Intensidad I_{xy}	15

3.4.2	Filtro de líneas <i>Lxy</i>	16
3.5	Mecanismo de control	22
3.6	Generación de probabilidades a priori.....	24
Capítulo 4:	Pruebas y resultados	27
4.1	Introducción	27
4.2	Medidas	28
4.2.1	Clase Pavimento.....	28
4.2.2	Clase Objeto.....	29
4.2.3	Clase Línea de carretera.....	30
4.2.4	Secuencias de análisis.....	31
4.3	Resultados	31
4.3.1	Clase pavimento	31
4.3.2	Clase objeto.....	33
4.3.3	Clase línea	36
4.3.4	Análisis subjetivo	39
Capítulo 5:	Conclusiones y trabajo futuro	43
Referencias	45

Índice de tablas

Figura 2. 1 Resultado de la transformación IPM	4
Figura 3.1 Diagrama de bloques del sistema	10
Figura 3.2 Histogramas y funciones estimadas	13
Figura 3.3 Proceso de obtención de pavimento.....	15
Figura 3.4 Aplicación del filtro de líneas.....	17
Figura 3.5 Mapas de probabilidad a priori para la clase objeto	24
Figura 4.1 Imagen original, imagen segmentada y segmentación de referencia .	27
Figura 4.2 Región correspondiente al vehículo segmentado en el plano IPM (en blanco), y líneas que describen la forma trapezoidal que la circunscribe.	29
Figura 4.3 Puntuaciones F-Score clase pavimento	32
Figura 4.4 Ángulos que parametrizan los laterales de los vehículos detectados en la secuencia 1	33
Figura 4.5 Ángulos que parametrizan los laterales de los vehículos detectados en la secuencia 2	33
Figura 4.6 Diferencias en píxeles entre el comienzo de los vehículos detectados y los de referencia	34
Figura 4.7 Segmentos de línea detectados en los tres sistemas y la imagen de referencia	36
Figura 4.8 Porcentaje de acierto para la detección de líneas en función de las filas eliminadas.....	37
Figura 4.9 Desviación media en grados de las líneas detectadas respecto de la referencia en función de las filas eliminadas	37
Figura 4. 10 Segmentos detectados para las 50 primeras filas de la imagen	38
Figura 4. 11 Imágenes de la secuencia 1 segmentadas por el algoritmo.....	39
Figura 4. 12 Imágenes de la secuencia 2 segmentadas por el algoritmo.....	40

Capítulo 1: Introducción

1.1 Motivación

Los sistemas de ayuda a la conducción (Advanced Driver Assistance Systems, ADAS) han supuesto un gran avance en materia de seguridad vial, desde la introducción de la frenada asistida en 1978 hasta los sistemas de conducción autónoma en la actualidad [1]. Gracias al aumento de potencia y reducción de coste de los microprocesadores, es posible implementar sistemas baratos basados en sensores que trabajen en tiempo real, capaces de modelar el entorno en el que se encuentra el vehículo para avisar al conductor en caso de peligro o incluso actuar de forma autónoma.

Estos sensores pueden ser activos o pasivos, entendiendo como activos aquellos que emiten señal para procesarla posteriormente y los pasivos los que no interfieren con el medio. Los sensores activos se dividen en tres clases: radar, laser y acústicos y presentan la ventaja frente a los pasivos de que son capaces de actuar bien en condiciones de niebla, lluvia o nieve mientras que con un gran número de vehículos moviéndose simultáneamente las señales pueden interferir. Los sensores pasivos, típicamente cámaras, si bien son más sensibles al medio permiten una implementación más sencilla y barata además de otorgar una descripción del entorno más completa.

En las últimas décadas se han realizado numerosos esfuerzos en investigación para la creación de algoritmos capaces de extraer información de alto nivel de las secuencias de vídeo provistas por cámaras integradas en vehículos, de forma que se le pueda dotar de capacidad para reaccionar ante el medio en situaciones de peligro para el conductor.

1.2 Objetivos

El objetivo de este Trabajo de Fin Grado es la implementación y análisis de un sistema capaz de segmentar el plano de carretera obtenido a partir de una transformación IPM en cuatro clases: pavimento, línea de carretera, vehículo y desconocido, donde esta última absorberá los elementos con baja probabilidad de pertenecer al resto de clases.

Dicho sistema se basa en el desarrollado en [3], si bien en este trabajo se han implementado cada una de las etapas proponiendo mejoras justificadas a algunas de ellas. Además, para evaluar de forma objetiva el funcionamiento del sistema y de las mejoras introducidas, se ha desarrollado un marco de pruebas específico para cada clase. Estas pruebas se basan en la comparación de diferentes métricas entre la imagen segmentada por el sistema y un conjunto de imágenes de referencia (*ground-truth*) que han sido creadas específicamente para este propósito. Finalmente estas pruebas son analizadas y se proponen una serie de pautas para el trabajo futuro

1.3 Organización de la memoria

La memoria presenta la siguiente estructura:

- ❖ Capítulo 1: Motivación y objetivos
- ❖ Capítulo 2: Estado del Arte
- ❖ Capítulo 3: Arquitectura del sistema implementado
- ❖ Capítulo 4: Pruebas y análisis de resultados
- ❖ Capítulo 5: Conclusiones y trabajo futuro

Capítulo 2: Estado del Arte

2.1 Introducción

El aumento de potencia de los microprocesadores y la introducción de sensores capaces de capturar vídeo digitalmente, ha permitido desarrollar numerosas aplicaciones para analizar y extraer información relevante de forma automática en diversos escenarios, uno de gran relevancia y en el que se centra este Trabajo de Fin de Grado es la automoción.

Podemos distinguir a grandes rasgos dos aproximaciones diferentes: por un lado están los trabajos basados en cámara fija (cámara estática situada en una posición fija monitorizando el entorno), enfocados principalmente a la detección de accidentes en carretera, maniobras de conducción ilegales o conteo de vehículos. Un ejemplo es [19], donde se propone una técnica que parte de una segmentación frente-fondo adaptativa, tras la cual se detectan y cuentan los vehículos mediante técnicas de seguimiento

En el otro están las aplicaciones donde la cámara se sitúa en el vehículo para dotarle de capacidad de reacción ante el medio, pudiendo así detectar situaciones de cambio de carril, colisiones inminentes o en última instancia permitir una conducción autónoma. Este escenario es más complejo ya que todos los elementos de la imagen están en constante movimiento, lo que exige robustez a los algoritmos ante posibles cambios de iluminación y entrada de nuevos elementos en la escena. En este entorno existen diferentes aproximaciones, clasificándose en base al número de cámaras usado: algunos autores como [9] o [11] usan dos cámaras para generar imágenes estéreo de la escena, pudiendo extraer información de profundidad y distancia a los objetos con mayor facilidad. Otros trabajos, como [8] o [3] usan sistemas mono-cámara que presentan la ventaja de ser más sencillos y de no necesitar calibración.

Se puede hacer otra clasificación en función de las regiones de interés a detectar. Así, hay trabajos que se encargan de modelar totalmente el entorno del vehículo [2] etiquetando de forma semántica la escena incluyendo peatones, líneas de carretera, vehículos, árboles... Otros trabajos se centran en detectar de forma precisa un único elemento, como pueden ser las líneas de carretera [5] [6] o vehículos [11]. En el presente trabajo, se implementa un sistema que detecta tres clases diferentes de elementos: pavimento, líneas y vehículos. Para ello se trabaja en un dominio proveniente de una transformación que elimina los efectos de perspectiva llamada Transformada de Perspectiva Inversa (*Inverse Perspective Mapping*, IPM). Esta transformada se ha utilizado en numerosos trabajos por las ventajas que presenta en la detección de líneas de carretera y modelado de carriles, por ello repasaremos primero ésta para posteriormente estudiar las técnicas más representativas de detección de vehículos y líneas de carretera.

2.2 Transformada de Perspectiva Inversa

La transformada de Perspectiva Inversa o IPM de ahora en adelante, fue introducida por Massimo Bertozzi y Alberto Broggi en [9]. Es una transformada geométrica diseñada para eliminar los efectos de perspectiva inherentes a una cámara situada en el frontal de un vehículo, que da como resultado una visión cenital de la escena (Figura 2.1).

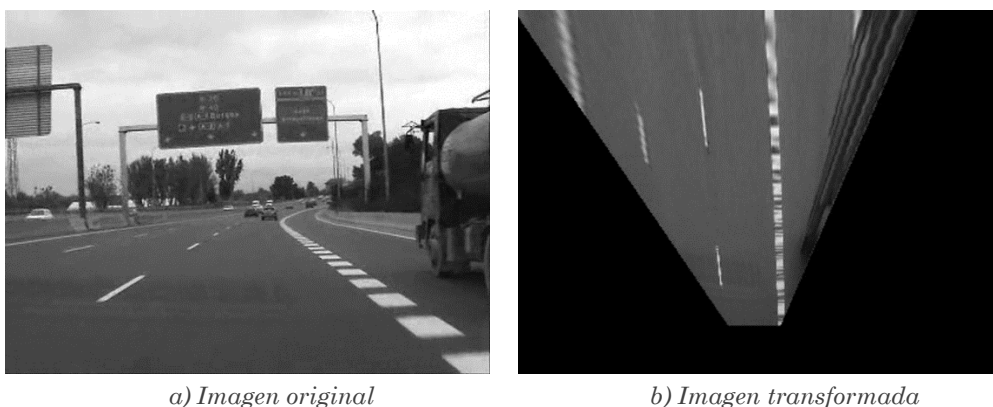


Figura 2. 1 Resultado de la transformación IPM

Esta transformada presenta una serie de ventajas que la hacen de especial utilidad en los sistemas de detección basados en visión. Gracias a que no se muestrea de forma homogénea la imagen de entrada para generar la transformada [13], las líneas de carretera presentan un ancho constante además de ser paralelas, propiedades muy útiles en la detección y modelado de líneas de carril. A su vez elimina uno de los efectos de la perspectiva, por el cual el mismo número de píxeles en la imagen original no representa el mismo espacio en el dominio real, haciendo que el cálculo de la velocidad y posición del vehículo sea más sencillo. Por contrapartida, cualquier elemento que tenga una altura mayor a la de la carretera se verá deformado, por lo que los vehículos pierden su forma y propiedades de simetría en este dominio.

Originalmente esta transformación requería saber las condiciones específicas de adquisición: posición de la cámara, orientación, etc. además de asumir que las carreteras eran totalmente planas. Posteriores trabajos relajaron esos requisitos, automatizando parte del proceso de calibración a partir del análisis de las imágenes entrada. Así por ejemplo en [14] aparte de introducir mejoras de trabajos anteriores que extraen información sobre la posición y la orientación de la cámara, se propone una forma de estimación del horizonte para eliminar los efectos producidos por pendientes ascendentes y descendentes, las cuales si no se tienen en cuenta resultan en carriles donde la asunción de paralelismo entre las líneas que los delimitan se pierde.

2.3 Detección de líneas de carril

La detección y modelado de líneas de carretera juega un papel fundamental en los sistemas de ayuda a la conducción. Gracias a estas técnicas el vehículo tiene la capacidad de saber por dónde ha de circular, encontrando aplicación en coches autónomos y seguridad, donde si existe un cambio de carril involuntario el vehículo puede actuar en consecuencia.

En [4] se define la detección de líneas de carril como el problema de localizar los límites del carril sin conocimiento previo de la geometría de la carretera. Tras la obtención de las líneas, es posible generar un modelo que las describa para poder hacer un seguimiento de éstas cuadro a cuadro. Puesto que en este Trabajo de Fin de Grado no se realiza un modelado de las líneas de carril, nos centraremos en repasar las técnicas existentes y más representativas para la detección.

De una forma general, existen dos tipos de aproximaciones diferentes a la detección, las basadas en patrones y las basadas en características [5].

La detección basada en patrones es más robusta frente al entorno, pues no depende tanto del estado de las líneas de carretera y de las condiciones de iluminación, pero poseen una falta de flexibilidad ante nuevas formas no contempladas en el diseño. En [6] se extraen características de la imagen para compáralas de forma estadística con la patrones y generar a partir del conjunto las líneas de carril un modelo matemático que las describa, obteniendo buenos resultados incluso en carreteras con líneas desgastadas.

Para la obtención basada en características muchos trabajos se basan en el uso de un detector de bordes, sistema que aunque simple, se ve muy afectado por cambios de iluminación o ambientales, unido a que presentan dificultades encontrando un umbral en intensidad para excluir aquellas componentes que no pertenezcan a las líneas [7].

Otros trabajos realizan la detección mediante análisis del histograma [8] donde la imagen se escanea por conjuntos de líneas. Haciendo asunciones sobre la forma esperada del histograma, concretamente que los píxeles con mayor de intensidad pertenecerán a las líneas, se obtienen aquellas regiones probables de ser línea. Tras esto, para filtrar aquellos valores que puedan pertenecer a otros elementos se aplica un árbol de decisiones, comparando con características de la zona inferior de la imagen, supuestamente formada únicamente por pavimento y líneas.

Finalmente, trabajos como [9] donde se introdujo la transformada IPM, asumen que en el dominio transformado las rectas tienen un ancho constante e introducen el uso de filtros morfológicos, los cuales tienen una respuesta alta a las líneas y se diseñan específicamente para éstas, siendo necesario conocer *a priori* características como su ancho o forma.

El artículo en el que se basa este Trabajo de Fin de Grado [3] hace uso de dos características diferentes para la obtención de las líneas, filtros morfológicos y el análisis del histograma.

2.4 Detección de vehículos.

La detección de vehículos basada en visión generalmente consta de dos etapas: una primera fase donde se recorre la imagen para obtener potenciales regiones de contener vehículos llamada generación de hipótesis, y una segunda conocida como verificación, donde se analizan en detalle los elementos contenidos en dichas zonas para determinar la existencia final de vehículos [1]. El presente trabajo permite la etapa de generación de hipótesis, de forma que se centrará el análisis en las principales técnicas presentes en el estado del arte, sin incluir aquellas de verificación.

A grandes rasgos los métodos de generación de hipótesis se pueden agrupar en tres clases: basados en visión estéreo, basados en movimiento y basados en conocimiento [10] [1].

Los métodos basados en visión estéreo explotan la capacidad de detectar profundidad gracias a la diferencia entre imágenes captadas por cada cámara (disparidad), así estas estrategias comparten la idea de buscar objetos volumétricos en el espacio 3D [11]. Presenta las ventajas de que se puede extraer información métrica de los objetos de forma sencilla, si bien requiere de una calibración y sincronización entre cámaras.

Por otro lado, los basados en movimiento obtienen las hipótesis a partir del análisis de los vectores de movimiento entre dos o más imágenes consecutivas. Los vehículos con una velocidad mayor al sujeto presentarán un patrón convergente, mientras que si llevan una velocidad menor será divergente. Un inconveniente asociado a estas técnicas son los altos recursos computacionales se necesitan, por eso en [20] los vectores de movimiento se calculan únicamente sobre la imagen de bordes.

Finalmente, los métodos basados en conocimiento son aquellos que hacen uso de información *a priori* para localizar posibles vehículos en la imagen. Una característica típica de un vehículo es que su parte trasera presenta simetría respecto de su eje vertical si lo situamos en el centro de este. Trabajos como [15] han explotado esta característica, donde la simetría se calcula tanto en el dominio original como en el de la imagen de bordes para reducir los efectos de brillos laterales en los vehículos que afectan a la simetría en la imagen de intensidad.

Otros trabajos usan características como los bordes [16], donde se les impone una serie de restricciones en forma, dirección y relación entre ellos que constituye conocimiento *a priori* sobre cómo es un vehículo. También existen aproximaciones basadas en el color, el cual es un elemento distintivo al presentar una cromaticidad mayor que el pavimento o las líneas.

Otro tipo de información *a priori* y que se utiliza en este Trabajo de Fin de Grado son las sombras. Caracterizadas por tener una intensidad reducida respecto del pavimento, ofrecen una característica robusta que se mantiene en la gran mayoría de situaciones. Este concepto fue introducido en [12] y algunos trabajos como [17] lo usan como información base de un clasificador que empareja sombras con vehículos, o [18], donde se usa esta característica entre otras como base para la detección de vehículos.

Capítulo 3: Arquitectura del sistema

3.1 Introducción

El objetivo de este sistema es segmentar las imágenes de un flujo de vídeo, cuadro a cuadro, proveniente de una cámara situada en la parte delantera de un vehículo mirando en la dirección de desplazamiento, en cuatro clases:

- ❖ Pavimento: zonas grises de bajo gradiente.
- ❖ Objetos: elementos de interés (vehículos) caracterizados por la sombra que proyectan en el pavimento.
- ❖ Línea: zonas de alta intensidad lumínica conformada por líneas o rectángulos.
- ❖ Desconocido: absorbe los demás elementos no considerados en otras clases y que arrojan una baja probabilidad de pertenecer a cualquiera de las clases anteriores.

Se parte de imágenes transformadas en el dominio IPM, explicado en capítulo 2.2. De esta forma se asume que los elementos que componen la imagen están restringidos a aquellos que podemos encontrar en un escenario típico de carretera con una observación cenital: pavimento, líneas de carril y vehículos.

Con esta información podemos hacer asunciones sobre cada imagen a segmentar, que constituyen el conocimiento *a priori*. Por un lado los objetos se caracterizan en base a su sombra, (lo que los hace independientes del color del propio objeto) y cuya principal característica es que presentan unos niveles medios de intensidad reducidos respecto al resto de elementos en la imagen. El pavimento vendrá caracterizado por ser una zona homogénea con pocas variaciones de intensidad entre los píxeles vecinos y finalmente las líneas aparecerán como rectángulos con niveles de intensidad altos rodeados normalmente por zonas de pavimento. Además puesto que se trabaja en el dominio transformado IPM las líneas de carretera presentan un ancho aproximadamente constante a lo largo de toda la imagen.

El sistema utiliza toda esta información para estimar distribuciones de verosimilitud representativas de los elementos existentes en la imagen IPM, las cuales serán optimizadas mediante el algoritmo *Expectation-Maximization* (EM), para en base a ellas realizar la segmentación bajo el marco de la teoría de decisión bayesiana.

La Figura 3.1 presenta un diagrama de bloques del sistema implementado. Inicialmente se procesa la imagen IPM en dos bloques para realizar una estimación de las regiones pertenecientes al pavimento y de las líneas de carretera. Las imágenes resultantes se usarán para extraer las poblaciones del resto de clases y modelar con ellas las funciones de verosimilitud que representen la probabilidad de pertenecer a ellas. Estas funciones se validan mediante un mecanismo de control que verifica que sus parámetros son correctos y que otorga robustez ante cambios repentinos de iluminación. Finalmente se utiliza la información del cuadro segmentado en el instante anterior y las funciones de verosimilitud estimadas para clasificar la imagen del instante de tiempo actual.

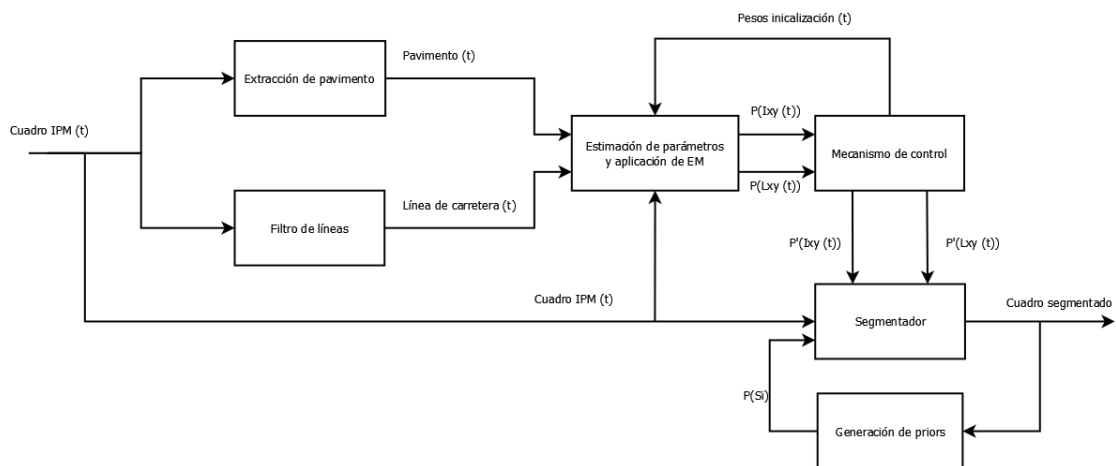


Figura 3.1 Diagrama de bloques del sistema

3.2 Decisión bayesiana

Usando la teoría de decisión bayesiana, la clasificación se realiza mediante la clase que maximice la probabilidad condicional a posteriori $P(S_i|\mathbf{z}_{xy})$, que se descompone mediante la regla de Bayes como:

$$P(S_i|\mathbf{z}_{xy}) = \frac{P(\mathbf{z}_{xy}|S_i)P(S_i)}{\sum_{i \in S} P(\mathbf{z}_{xy}|S_i)P(S_i)}$$

Denotemos por $S = \{S_p, S_o, S_l, S_u\}$ al conjunto de clases que representan respectivamente el pavimento, los objetos, las líneas y los elementos desconocidos (*unknown*).

\mathbf{z}_{xy} es el vector de observaciones asociadas a un píxel con coordenadas (x, y) . Llamamos $P(\mathbf{z}_{xy}|S_i)$ a la verosimilitud de S_i respecto a \mathbf{z}_{xy} , que representa, dada una clase S_i , la probabilidad de que un píxel presente un vector de observaciones asociado \mathbf{z}_{xy} .

$P(S_i)$ representa la probabilidad con la que un píxel pertenece a la clase S_i antes de realizar alguna observación sobre él, es decir, la probabilidad *a priori*, sujeta a la restricción:

$$\sum_{i \in S} P(S_i) = 1.$$

Esta probabilidad *a priori* se extrae de la clasificación en el instante de tiempo anterior en el bloque *Generación de Priors* de la Figura 3.1, aprovechando la redundancia temporal del vídeo en la que cada elemento entre cuadros se desplaza un número reducido de píxeles.

Finalmente, la evidencia: $\sum_{i \in S} P(\mathbf{z}_{xy}|S_i)P(S_i)$ es un factor de escala que asegura que las probabilidades resultantes estarán entre 0 y 1.

La fórmula de Bayes combina estos factores para obtener la distribución a posteriori $P(S_i|\mathbf{z}_{xy})$. De la misma forma que $P(S_i)$ nos da la probabilidad de pertenecer a una clase antes de la observación, $P(S_i|\mathbf{z}_{xy})$ refleja las probabilidades de pertenecer a una clase S_i tras las observaciones.

La regla de decisión de Bayes dicta que aquella clasificación que minimiza la probabilidad de error es aquella que maximice la probabilidad a posteriori, de esta forma asociaremos el píxel a la clase S_i si maximiza la probabilidad $P(S_i|\mathbf{z}_{xy})$.

Esta clasificación se realiza en el bloque *Segmentador*, mostrado en la Figura 3.1. Como se observa, este tiene tres entradas:

- ❖ $P(I_{xy}|\{S_i\}_{i \in S})$: función de verosimilitud que modela, para cada clase, para cada clase de los distintos valores de intensidad.
- ❖ $P(L_{xy}|\{S_i\}_{i \in S})$: función de verosimilitud que modela, para cada clase, la respuesta a un filtro de líneas
- ❖ $P(S_i)$: Matriz de probabilidades *a priori*.

I_{xy} y L_{xy} representan dos características a nivel de píxel: su nivel de intensidad y su respuesta a un filtro de líneas respectivamente. Este filtro se ha diseñado para dar alta respuesta a aquellos píxeles pertenecientes a las líneas de carretera aprovechando la morfología que presentan en el dominio IPM.

Para la clasificación, en este trabajo se asume por simplicidad que I_{xy} y L_{xy} se pueden considerar estadísticamente independientes, resultando $P(\mathbf{z}_{xy}|S_i) = P(I_{xy}|S_i) P(L_{xy}|S_i)$.

La salida será la imagen segmentada, una imagen donde el valor de los píxeles indica la pertenencia a una determinada clase.

3.3 Funciones de verosimilitud

Las funciones de verosimilitud utilizadas explotan el conocimiento *a priori* que se tiene sobre la imagen IPM. $P(I_{xy}|\{S_i\}_{i \in S})$ modela la probabilidad de pertenecer a cada clase en función de la intensidad o nivel de gris, puesto que podemos asumir que las imágenes IPM en escala de grises, presentan un histograma tri-modal como muestra la Figura 3.2, debido a la restricción de elementos que las componen, la parte inferior del histograma corresponderá a los elementos oscuros pertenecientes a la sombra de los vehículos; la superior, con menos píxeles, a las líneas de carretera; y la central, con valores de intensidad medios, al pavimento. Por tanto, podemos modelar el histograma como una mezcla de tres gaussianas, siendo la inferior la representativa de la clase vehículo, la central de la clase pavimento y la superior de la clase línea.

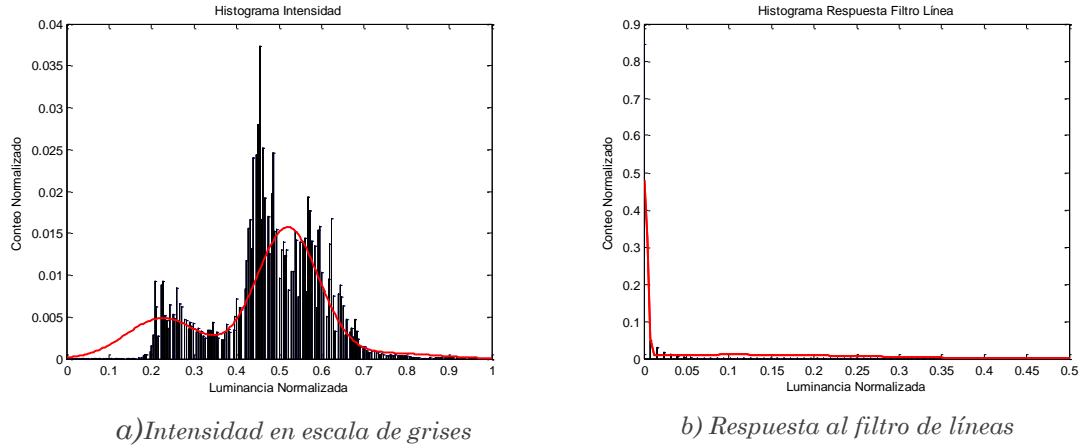


Figura 3.2 Histogramas y funciones estimadas

La segunda función $P(L_{xy}|\{S_i\}_{i \in S})$, modela la probabilidad de pertenencia a cada clase en función de la respuesta a un filtro de líneas que muestra una alta respuesta para los píxeles pertenecientes a las líneas de carretera y baja al resto de elementos de la imagen. Una imagen filtrada, como se explica en la sección 3.3.2 presenta un histograma bi-modal, modelable como la mezcla de dos gaussianas, una centrada en cero correspondiente a la respuesta de los objetos y el pavimento, y otra con mayor varianza cubriendo la cola del histograma representativa de la respuesta a las líneas.

Formalizando las expresiones, ambas son funciones paramétricas correspondientes a una mezcla de gaussianas unidimensional, la cual queda definida por los siguientes parámetros:

- ❖ k : número de gaussianas
- ❖ μ_i : vector de medias.
- ❖ σ_i : vector de varianzas.
- ❖ w_i : vector de pesos, representan el grado con el que cada gaussiana contribuye a la distribución total.

Por tanto,

$$P(I_{xy}|\{S_i\}_{i \in S}) =: \sum_{i \in S} w_{i,I} P(I_{xy}|S_i) \quad P(L_{xy}|\{S_i\}_{i \in S}) =: \sum_{i \in S} w_{i,L} P(L_{xy}|S_i)$$

donde:

$$P(I_{xy}|S_i) =: \exp\left(-\frac{1}{2\sigma_{I,i}^2}(I_{xy} - \mu_{I,i})^2\right) \quad P(L_{xy}|S_i) =: \exp\left(-\frac{1}{2\sigma_{L,i}^2}(L_{xy} - \mu_{L,i})^2\right)$$

De esta forma para caracterizar la función $P(I_{xy}|\{S_i\}_{i \in S})$ es necesario estimar los parámetros $\{\mu_{Ip}, \mu_{Io}, \mu_{Il}\} \{\sigma_{Ip}, \sigma_{Io}, \sigma_{Il}\} \{w_{Ip}, w_{Io}, w_{Il}\}$.

A su vez, para $P(L_{xy}|\{S_i\}_{i \in S})$ los parámetros correspondientes a la clase objeto y pavimento no pueden ser distinguidos, teniendo los mismos valores. Por tanto los parámetros a estimar son $\{\mu_{Lp} = \mu_{Lo}, \mu_{Ll}\} \{\sigma_{Lp} = \sigma_{Lo}, \sigma_{Ll}\} \{w_{Lp}, w_{Lo}, w_{Ll}\}$

Finalmente, la clase desconocido modelaría la probabilidad de pertenecer a otros elementos de la escena, lo que se asume igual de probable para los diferentes valores de intensidad y respuesta de la imagen al filtro de líneas. De esta forma su estimación no se realiza y se parametriza como una gaussiana centrada a mitad del histograma y con gran varianza para que se aproxime a una distribución uniforme.

3.4 Estimación de parámetros

La estimación de los parámetros se realiza en el bloque *Estimación de Parámetros y Aplicación del Algoritmo EM* de la Figura 3.1. Primero, para cada función de verosimilitud se obtiene una primera aproximación de los parámetros μ_i y σ_i para optimizarlos posteriormente mediante el algoritmo EM donde se calcularán los pesos asociados a cada distribución w_i .

3.4.1 Intensidad I_{xy}

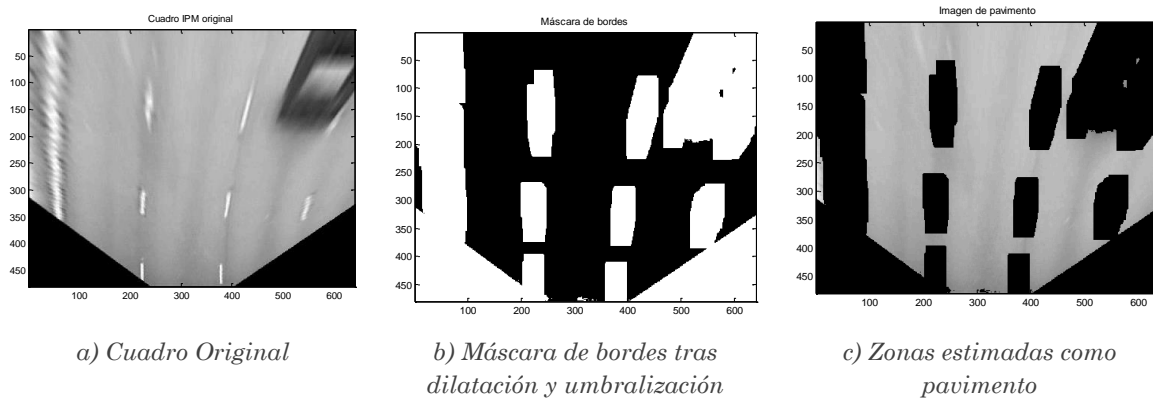


Figura 3.3 Proceso de obtención de pavimento

La obtención de los parámetros que caracterizan la función $P(I_{xy}|\{S_i\}_{i \in S})$ parte de una caracterización de los píxeles que pertenecen al pavimento. Esto se realiza por dos razones: el pavimento compone una parte muy significativa del plano de carretera en la imagen transformada IPM ofreciendo así datos suficientes para generar estadísticos robustos y fiables. Además al ser una zona eminentemente homogénea, la determinación de regiones pertenecientes a dicha clase es relativamente sencilla mediante la eliminación de zonas que presenten cambios bruscos en intensidad (bordes). Finalmente, puesto que el histograma de intensidad se asume tri-modal por la naturaleza de los elementos que componen la imagen IPM, si se realiza una correcta estimación de los píxeles que componen el pavimento, se puede umbralizar la imagen para determinar las regiones con niveles de intensidad con una gran desviación respecto del pavimento, obteniendo así las regiones representativas de los objetos y las líneas.

La estimación de las regiones pertenecientes al pavimento se realiza en el bloque *Extracción de Pavimento* y como muestra la Figura 3.3 se actúa en 3 pasos: Primero se aplica un detector de bordes a la imagen para distinguir entre las zonas homogéneas y aquellas que presentan transiciones, concretamente el sistema implementado utiliza el detector de Canny [19]. Posteriormente la imagen de bordes es dilatada mediante operadores morfológicos [22] para generar una máscara que elimine las zonas colindantes de los bordes, típicamente las líneas de carretera e interior de vehículos.

La imagen resultante de aplicar la máscara estará compuesta por zonas de pavimento y por las regiones interiores de las sombras que se proyecten en el pavimento que el operador morfológico no haya sido capaz de eliminar. Un umbral heurístico y que ha ofrecido buenos resultados durante las pruebas, es excluir aquellas regiones cuyo valor sea inferior a la desviación típica de los píxeles de la imagen tras la aplicación de la máscara. Finalmente las regiones estimadas de pavimento se ilustran en la Figura 3.3 c).

Sobre la imagen de pavimento estimada se obtienen los parámetros de la distribución que modela la probabilidad de pertenecer a la clase pavimento mediante el cálculo de la media μ_{Ip} y la desviación típica σ_{Ip} de los valores que la componen.

De nuevo, asumiendo que los elementos que conforman la imagen IPM serán fundamentalmente pavimento, líneas de carretera, vehículos y sus sombras, se umbraliza el cuadro original IPM a $\mu_{Ip} \pm 3\sigma_{Ip}$. De forma que se satisfacen dos premisas: que la relación entre medias cumplirá $\mu_{Io} < \mu_{Ip} < \mu_{Il}$ y que solo aquellos píxeles que caen fuera del pavimento con una probabilidad del 99,99% se usan para modelar el resto de clases [3]. Sobre las regiones resultantes de umbralizar la imagen se obtienen los parámetros de la clase objeto $\{\mu_{Io}, \mu_{Il}\}$ y línea $\{\sigma_{Io}, \sigma_{Il}\}$.

3.4.2 Filtro de líneas L_{xy}

Las líneas de carretera en el plano IPM poseen dos características que las hacen fácilmente identificables del resto de elementos, poseen un ancho constante y aparecen como pulsos de alta intensidad rodeados de zonas más oscuras. Aprovechando estas dos características en [3] se diseña un filtro de líneas usado para modelar de forma más precisa la pertenencia a la clase línea aprovechando el conocimiento *a priori* que se tiene sobre su morfología y evitar así clasificaciones de elementos con alta intensidad que no son realmente línea. La expresión del filtro es:

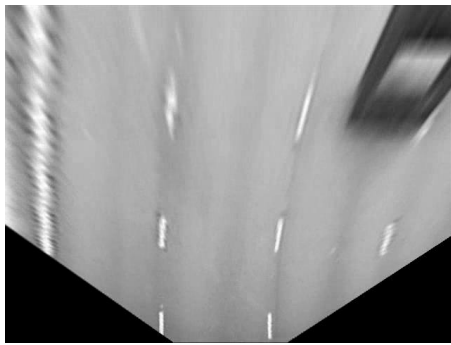
$$y_i = 2x_i - (x_{i-\tau} + x_{i+\tau}) - |x_{i-\tau} - x_{i+\tau}|$$

donde y_i representa el píxel de la fila i en la imagen filtrada, x_i el píxel de la fila i de la imagen IPM original y τ el ancho estimado de la línea en píxeles.

La idea general del filtro es la siguiente: para cada píxel de la fila escogida en la imagen, obtener su diferencia con los aquellos píxeles separados τ a izquierda y derecha. De esta manera, si estamos en una zona de alta intensidad correspondiente a la línea, la diferencia entre el píxel de la línea y aquellos separados su ancho (τ) arrojará un valor alto y positivo. Sin embargo si estamos

en zonas homogéneas la comparación entre el píxel central y sus adyacentes será cercana a cero. A su vez, si estamos comparando zonas de pavimento u objetos cercanos a líneas la diferencia será alta, pero arrojará valores negativos.

El último término de la expresión está enfocado a penalizar los casos en los que la diferencia entre los píxeles situados τ a derecha y τ a izquierda sea alta, como por ejemplo las líneas situadas entre pavimento y vehículo; donde el valor de salida tendría un valor mayor que si la línea se encontrase ubicada en una zona homogénea.



a) Cuadro original



b) Imagen tras la aplicación del filtro de líneas

Figura 3.4 Aplicación del filtro de líneas

Para una imagen de carretera IPM la respuesta del filtro presentará un histograma bimodal, con un gran número de elementos cercanos a cero con poca varianza debido a las respuestas de los píxeles pertenecientes a pavimento y objeto, como muestra la Figura 3.4 b), y una segunda componente con mucha más varianza correspondiente a la respuesta de los píxeles contenidos en líneas de carretera los cuales en función del entorno ofrecen respuestas muy diferentes.

Finalmente, las dos componentes del histograma se separan utilizando la desviación típica de la imagen. Obteniendo $\{\mu_{Lp} = \mu_{Lo}, \sigma_{Lp} = \sigma_{Lo}\}$ con los píxeles cuyo valor es menor a la suma de la media y la desviación típica de los píxeles de la imagen filtrada y $\{\mu_{Ll}, \sigma_{Ll}\}$ con los píxeles cuyo valor es mayor al umbral anterior. Esta separación propuesta en [3] ha ofrecido buenos resultados durante las pruebas.

3.4.3 Algoritmo *Expectation-Maximization*

Una vez se ha obtenido una primera aproximación de los parámetros de las distribuciones, estos son optimizados a través del algoritmo *Expectation-Maximization*, donde además se calculará de forma iterativa los pesos asociados a cada gaussiana.

3.4.3.1 Marco teórico:

Partimos de un conjunto de observaciones o muestras asociadas a diferentes clases, donde las muestras son independientes e idénticamente distribuidas aleatoriamente. Asumimos que la probabilidad de que dada una clase S_i el conjunto de muestras Y pertenezca a ella, tiene una forma paramétrica conocida, determinada por un conjunto de parámetros θ_i : $p(Y|S_i, \theta_i)$. La cuestión por tanto es usar la información obtenida por las muestras observadas para realizar buenas estimaciones de los parámetros θ_i asociados a cada clase.

Llamamos $p(S_i|\theta_i) = \prod_{k=1}^n p(Y_k|\theta_i)$ a la función de verosimilitud de θ_i respecto al conjunto de muestras z . La estimación de máxima verosimilitud es por definición aquel valor que maximiza $p(S_i|\theta_i)$, es decir, aquel valor de θ_i que mejor define la distribución de probabilidad de los valores observados para dicha clase. Por tanto, el objetivo es hallar los parámetros de θ_i que maximizan dicha función.

El algoritmo EM, es un método iterativo para encontrar una estimación de máxima verosimilitud de los parámetros que definen una distribución de probabilidad cuando existen datos ocultos o erróneos. En el marco de este TFG existen tres elementos que contribuyen al ruido, el propio sensor de la cámara al captar la imagen, la transformada IPM, que al basarse en un mecanismo de interpolación añade ruido y finalmente la etapa de estimación de pavimento, la cual no es perfecta e introduce valores píxeles pertenecientes a otras clases.

Denotemos por θ al vector de parámetros que queremos estimar, por $Z = \{z_1, z_2, \dots, z_m\}$ al conjunto de m datos observados, por $X = \{x_1, x_2, \dots, x_n\}$ al conjunto de n datos no observados u ocultos y finalmente por $Y = Z \cup X$ a la unión de los conjuntos de datos observados y ocultos, es decir, el conjunto de datos completos.

X al ser desconocido se modela como una variable aleatoria dependiente del conjunto de datos observados Z , que a su vez depende de los parámetros θ a estimar, por lo que X depende de Z y θ . Y , por tanto es otra variable aleatoria al estar definida en términos de X .

Nuestro objetivo por tanto, es encontrar aquel vector θ que maximice $p(Y|\theta)$: la distribución de probabilidad que define Y y que depende de θ . Dicha distribución nos da la verosimilitud de los datos completos Y dado θ .

Debido a que es más sencillo calcular el θ que maximice la log verosimilitud, se le aplica el logaritmo resultando $\log p(Y|\theta)$. Debido a que el logaritmo es una función monótonamente creciente la solución será la misma [20].

El algoritmo EM busca la hipótesis θ' que maximice $p(Y|\theta')$, pero dado que realmente no conocemos Y , deberemos maximizar el valor esperado de $\log p(Y|\theta')$. Para lograrlo, EM actúa en los siguientes pasos:

Inicialización

Escoge un valor inicial para $\theta^{(m)}$

Paso *Expectation*

En este paso se calcula el valor esperado de $\log p(Y|\theta)$ respecto de $\theta^{(m)}$ y de los datos observados Z .

Para ello, dado el conjunto de datos observados Z y asumiendo que nuestra hipótesis $\theta^{(m)}$ es correcta, calculamos cómo de probable es que el conjunto de datos completos sea Y , es decir, calculamos $(Y|Z, \theta^{(m)})$.

Ahora para calcular el valor esperado de $\log p(Y|\theta)$, integraremos sobre el conjunto de posibles valores de Y , pesando cada posible valor de $\log p(Y|\theta)$ por la probabilidad que tiene ese valor Y dados θ y Z . Este valor esperado recibe el nombre de función Q y se expresa como:

$$Q(\theta, \theta^{(m)}) = E_{Y|Z, \theta^{(m)}}[\log p(Y|\theta)] = \int_{Y \in \psi} \log p(Y|\theta) p(Y|Z, \theta^{(m)}) dy$$

donde, $\theta^{(m)}$ es nuestra mejor estimación del vector de parámetros para la distribución. θ es la variable de la que depende Q y la candidata para una estimación mejorada, finalmente ψ es el espacio de valores que nuestros datos pueden tomar.

Paso *Maximization*

En este paso se calcula aquél valor de θ que maximiza $Q(\theta, \theta^{(m)})$ y se actualiza la hipótesis $\theta^{(m+1)}$ con dichos parámetros.

EM por tanto comienza el algoritmo con una hipótesis inicial sobre el vector de parámetros. Estima la función Q con dicha hipótesis y encuentra los valores de θ que la maximizan. Este proceso itera hasta que no se puede encontrar un valor de θ que maximice $Q(\theta, \theta^{(m)})$

3.4.3.2 EM para una mezcla de gaussianas:

En el problema que nos ocupa en este Trabajo de Fin de Grado los datos se han modelado como una mezcla de gaussianas, por tanto es necesario particularizar las expresiones anteriores en este tipo de distribución. El desarrollo detallado de las expresiones se encuentra en [20], mostrándose aquí se particularizadas para el caso unidimensional con k gaussianas con n muestras independientes e idénticamente distribuidas aleatoriamente denotadas por y .

Definimos:

1. Los parámetros de una mezcla de gaussianas $\theta_j = \{w_j, \mu_j, \sigma_j\}_{j=1}^k$ representando el peso, la media y la varianza de la distribución j respectivamente.
2. Distribución de probabilidad gaussiana parametrizada con media μ y desviación típica σ :

$$\Phi(x|\mu, \sigma) =: \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$$

3. La probabilidad de que la i -ésima muestra pertenezca a la j -ésima gaussiana como:

$$\gamma_{ij} = \frac{w_j \Phi(z_i|\mu_j, \sigma_j)}{\sum_{l=1}^k w_l \Phi(z_i|\mu_l, \sigma_l)}$$

Que satisface:

$$\sum_{i=1}^k \gamma_{ij} = 1$$

El algoritmo sigue los siguientes pasos:

Inicializar parámetros

Paso *Expectation*:

1. $\gamma_{ij}^{(m)} = \frac{w_j^{(m)} \Phi(z_i | \mu_j^{(m)}, \sigma_j^{(m)})}{\sum_{l=1}^k w_l^{(m)} \Phi(z_i | \mu_l^{(m)}, \sigma_l^{(m)})}, i = 1, \dots, n, j = 1, \dots, k$
2. $n_j^{(m)} = \sum_{i=1}^n \gamma_{ij}^{(m)}, j = 1, \dots, k. \quad (\sum_{i=1}^j n_j = n, j = 1, \dots, k).$

Paso *Maximization*:

3. $w_j^{(m+1)} = \frac{n_j^{(m)}}{n}, j = 1, \dots, k$
4. $\mu_j^{(m+1)} = \frac{1}{n_j} \sum_{i=1}^n \gamma_{ij}^{(m)} y_i, j = 1, \dots, k$
5. $\sigma_j^{2(m+1)} = \frac{1}{n_j} \sum_{i=1}^n \gamma_{ij}^{(m)} (y_i - \mu_j^{(m+1)})^2$

Comprobar convergencia

La inicialización de parámetros se realiza con las estimaciones de $\{\mu_{Ip}, \mu_{Io}, \mu_{Il}\}$, $\{\mu_{Lp}, \mu_{Lo}, \mu_{Ll}\}$, $\{\sigma_{Ip}, \sigma_{Io}, \sigma_{Il}\}$, $\{\sigma_{Lp}, \sigma_{Lo}, \sigma_{Ll}\}$ realizadas a partir de la obtención del pavimento y el filtro de líneas. La clase desconocido no se introduce dentro del algoritmo EM y se mantiene estática para todas las imágenes de entrada. Los pesos para la primera se inicializan con $\frac{1}{\text{Nº de clases}} = \frac{1}{4}$ y posteriormente son inicializados con el valor obtenido en la iteración anterior.

La ejecución de los pasos E-M intuitivamente puede verse de la siguiente manera,

Para el paso *Expectation*:

- ❖ En 1. se calcula la probabilidad de cada muestra i de pertenecer a cada distribución j ($w_j^{(m)} \Phi(z_i | \mu_j^{(m)}, \sigma_j^{(m)})$) y se pesa por la suma total de probabilidades para que el total de la distribución sume 1.
- ❖ Posteriormente en 2. se estima la contribución de cada gaussiana al total mediante la suma de las distribuciones por separado.

Los resultados obtenidos en este paso no suponen un cálculo de las esperanza de las distribuciones si bien en [20], el autor las sitúa ahí para mantener la correspondencia con el algoritmo original, pero podrían haberse introducido directamente en el Paso *Maximization*.

Paso *Maximization*:

En este paso se realiza el proceso de cálculo de la esperanza y maximización a la vez [21].

- ❖ En 3. se obtienen los nuevos pesos de cada distribución, calculados como la probabilidad de que todo el conjunto de muestras pertenezca a la distribución j entre el número total de datos
- ❖ En 4. Se obtienen las medias que maximizan la esperanza, parámetro a optimizar en la log-verosimilitud para una mezcla de gaussianas. Puesto que para una distribución gaussiana se cumple que la esperanza de la distribución es igual a la media, si se maximizan las medias se maximiza la esperanza.
- ❖ Finalmente en 5. Se obtiene la varianza, la cual nos da una idea de cómo de dispersos están los datos respecto a la media.

El algoritmo itera hasta que converja. La forma elegida en este caso es que el valor absoluto de la log-similitud entre dos iteraciones consecutivas sea inferior a un umbral predefinido.

3.5 Mecanismo de control

Este bloque, como muestra la Figura 3.1 tiene como entradas los parámetros que definen las distribuciones de intensidad y filtro de líneas calculadas en la iteración actual y almacena los últimos parámetros que se determinaron como válidos. Su salida se compone por los parámetros de las distribuciones que se usarán para clasificar los elementos del cuadro actual y los pesos con los que se inicializará el algoritmo EM en la siguiente iteración.

Aquí se detectan y corrigen posibles situaciones que comprometan las estimaciones del sistema. Estos casos producen que las asunciones que se hacen sobre la naturaleza de las imágenes IPM de carretera no sean correctas y por tanto se hagan clasificaciones erróneas. Típicamente son cambios muy significativos entre cuadros consecutivos a los que el sistema no puede adaptarse, como la entrada en túneles o cambios en la naturaleza del pavimento que producen como resultado inicializaciones incorrectas de las etapas anteriores como la estimación de pavimento o filtro de líneas.

EM puesto que es un algoritmo muy sensible a las inicializaciones [20] también producirá salidas incorrectas. Para evitar que estas situaciones comprometan la robustez del sistema se introduce un mecanismo de control sencillo que actúa en dos niveles.

En el primer nivel se comprueba los parámetros que conforman ambas distribuciones tras el algoritmo EM $P(I_{xy}|\{S_i\}_{i \in S})$ y $P(L_{xy}|\{S_i\}_{i \in S})$ sean válidos y coherentes con el rango de intensidades de la imagen.

Si alguno de los parámetros no cumple los requerimientos del primer nivel, el sistema entra en un estado transitorio donde la salida la conforman los últimos parámetros determinados como válidos y los pesos se reestablecen a su valor de inicialización original. Este nivel está diseñado para detectar las situaciones donde, por las razones expuestas anteriormente, el sistema falle de un cuadro a otro y así evitar que se generen probabilidades *a priori* y pesos de inicialización incorrectos que produzcan clasificaciones posteriores incorrectas.

Además se introduce una propuesta sencilla enfocada a cumplir dos propósitos: evitar que sucesivas malas estimaciones que cumplan los requisitos del primer nivel degeneren el sistema, y suavizar la adaptación teniendo en cuenta los parámetros de las distribuciones anteriores en el tiempo. Para ello en este segundo nivel, se realiza la media ponderada de los vectores media y varianza que definen la función $P(I_{xy}|\{S_i\}_{i \in S})$ con aquellos del instante anterior, De esta forma se mantiene un compromiso, permitiendo al sistema que evolucione en función del entorno, pero de forma controlada.

$$\mu_t = \alpha * \mu_t + (1 - \alpha) * \mu_{t-1} \qquad \sigma_t = \alpha * \sigma_t + (1 - \alpha) * \sigma_{t-1}$$

Donde μ_t, σ_t representan respectivamente el vector de medias y varianzas del instante actual y μ_{t-1}, σ_{t-1} del instante anterior, finalmente α es el factor de ponderación y está acotado entre cero y uno.

Los parámetros de $P(L_{xy}|\{S_i\}_{i \in S})$ no se introducen en este mecanismo debido a que estos sí que dependen mucho de la estimación en el cuadro actual puesto que el filtro de líneas es muy sensible al entorno y su respuesta presentará variaciones importantes entre imágenes consecutivas. Además si se trabaja con secuencias de vídeo IPM estabilizadas en las que la transformada depende del horizonte en cada imagen, al tener diferentes homografías entre cuadros las líneas variarán ligeramente de ancho, lo que ofrece de nuevo respuestas diferentes.

3.6 Generación de probabilidades a priori

Debido a la alta redundancia temporal que existe en el vídeo, mediante un análisis de este se puede extraer información útil para las clasificaciones en el instante de tiempo actual. En este sistema se utilizan las clasificaciones del cuadro anterior para obtener las probabilidades *a priori* $P(S_i)$, las cuales representan el conocimiento previo sobre la clase a la que pertenece un pixel con coordenadas (x, y) .

Se puede asumir que un elemento de la imagen no se desplaza un gran número de píxeles entre dos cuadros consecutivos, una primera aproximación por tanto sería dar valores altos a $P(S_i)$ para aquellos píxeles clasificados como pertenecientes a S_i y bajos para el resto, de forma que se cumpla $\sum_{i \in S} P(S_i) = 1$.

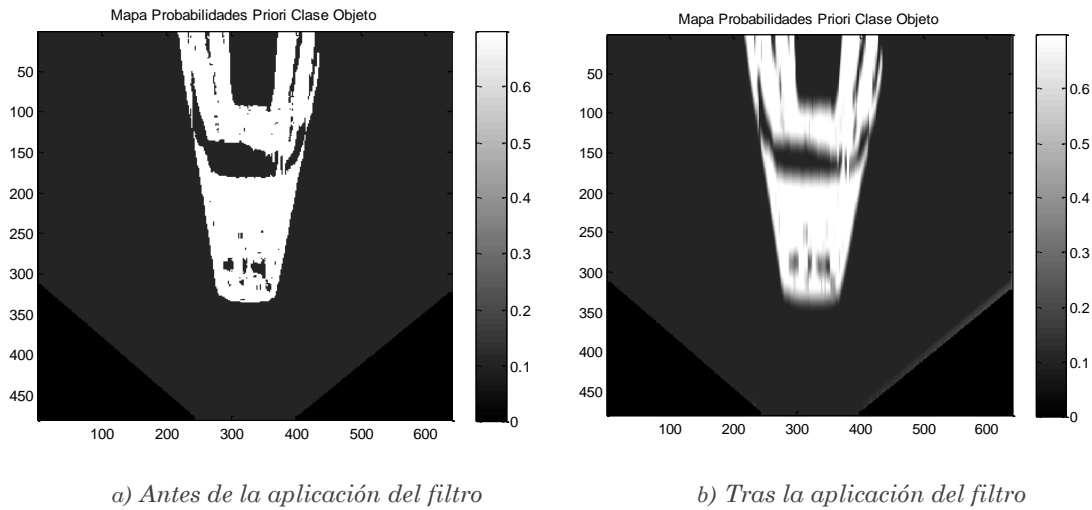


Figura 3.5 Mapas de probabilidad *a priori* para la clase objeto

No obstante, debido a que la fuente de vídeo proviene de un vehículo desplazándose normalmente a alta velocidad y en línea recta, se aplica un posprocesado a la matriz de probabilidades *a priori* para simular la incertidumbre con la que un elemento se desplaza entre cuadros, como se muestra en la Figura 3.5. Para este propósito se genera una máscara gaussiana vertical de tamaño $[n, 1]^T$ donde n representa la longitud del vector en píxeles, asociada al movimiento del vehículo portador de la cámara. Si el sistema contase con información sobre la velocidad se podrían adaptar los parámetros para generar máscaras variables de longitud y desviación típica pero puesto que no se dispone de ella se ha decidido dejar los parámetros fijos.

Adicionalmente, durante las pruebas del sistema se detectó un problema asociado a aplicar la máscara en los píxeles de los límites imagen, donde la suma de sus elementos no es 1 y por tanto la suma de probabilidades *a priori* entre clases tampoco. Para solucionarlo se decidió replicar las filas superiores e inferiores el número de píxeles necesario para que la suma sea 1 y no generar probabilidades incorrectas (*improper priors*).

Capítulo 4: Pruebas y resultados

4.1 Introducción

El objetivo de estas pruebas es realizar una evaluación objetiva del sistema descrito en el capítulo 3. Se compararán tres versiones del sistema: una simplificada sin algoritmo EM ni mecanismo de control, una segunda que incluye el algoritmo EM y el mecanismo de control de primer nivel y una tercera con todas las funcionalidades implementadas, incluyendo el mecanismo de control adaptativo (segundo nivel) propuesto.

Las diferentes medidas se realizarán entre imágenes IPM segmentadas por el sistema e imágenes de referencia (*ground-truth*) de las segmentaciones que han sido etiquetadas a mano. Estas se ilustran en la Figura 4.1, donde las etiquetas resultantes de segmentar la imagen se representan mediante niveles de intensidad diferentes.

Finalmente se muestra un set de imágenes segmentadas por el algoritmo para evaluar de forma subjetiva su rendimiento

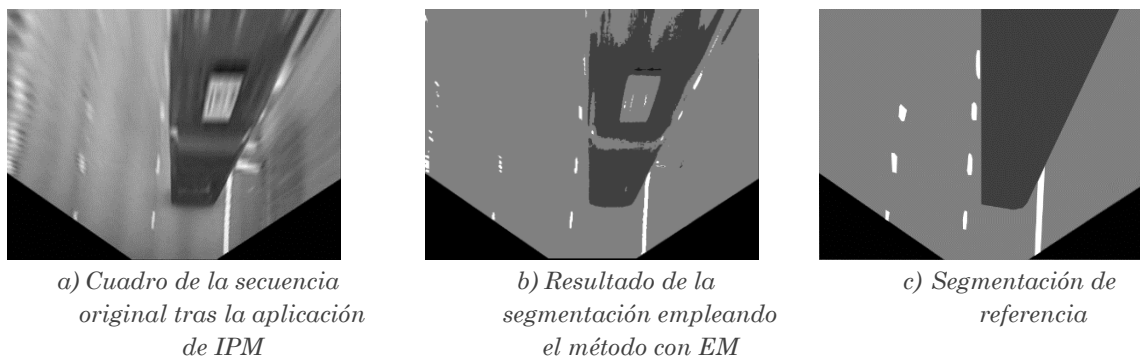


Figura 4.1 Imagen original, imagen segmentada y segmentación de referencia

4.2 Medidas

Puesto que la detección de los elementos de cada clase presenta un interés diferente, se aplicarán diferentes medidas de evaluación a cada una de ellas que se detallan a continuación:

4.2.1 Clase Pavimento

Esta clase no presenta ninguna forma geométrica definida, de este modo su evaluación se centrará en obtener medidas basadas en el número de píxeles correctamente clasificados. Para este propósito usaremos el indicador *F-Score*, el cual se expresa en función de la precisión (*precision*) y el nivel de acierto (*recall*) que a su vez se definen en función de TP, FP y FN.

Definimos:

- ❖ TP o *True Positive*: píxeles clasificados correctamente como pertenecientes a una clase S .
- ❖ FP o *False Positive*: píxeles clasificados como pertenecientes a la clase S pero que en realidad no pertenecen a dicha clase.
- ❖ FN o *False Negative*: píxeles clasificados como no pertenecientes a S que deberían haber sido etiquetados como pertenecientes.
- ❖ *Precision*: para una clase S , el cociente entre el número total de píxeles que han sido clasificados correctamente y el total de píxeles clasificados como pertenecientes a S . Se expresa en función de TP y FP.

$$Precision(S) = \frac{TP}{TP + FP}$$

- ❖ *Recall*: es un estimador que indica, de todos los píxeles que deberían haber sido clasificados como pertenecientes a una determinada clase S , cuántos fueron correctamente clasificados. Se expresa en función de TP y FN.

$$Recall(S) = \frac{TP}{TP + FN}$$

❖ *F-Score*: media armónica entre *precision* y *recall*.

$$Fscore = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 * Precision * Recall}{Precision + Recall}$$

Dado que *precision* y *recall* están acotados entre 0 y 1, donde valores cercanos a 1 indican que el número de falsas clasificaciones (FP o FN) es muy bajo en comparación a las clasificaciones correctas (TP), *F-Score* al ser un promediado aritmético de ambos también está entre estos valores, donde valores cercanos a 1 expresan que tanto *precision* como *recall* son cercanos a 1.

4.2.2 Clase Objeto

Un vehículo en el plano IPM debido a la transformación realizada presenta generalmente una forma trapezoidal. El origen de esta forma (límite inferior), marcado por la sombra normalmente es paralelo al borde inferior de la imagen, mientras que los bordes laterales tienden a alejarse hacia los extremos de la imagen. Así, el vehículo queda determinado por tres rectas: una horizontal que marca el comienzo del vehículo y otras dos que determinan sus límites laterales. Un caso típico es el que ilustra la Figura 4.2.

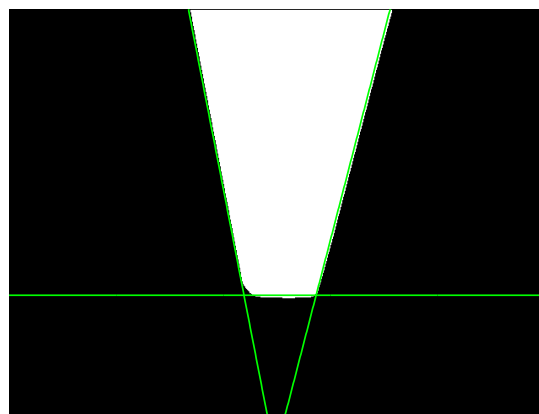


Figura 4.2 Región correspondiente al vehículo segmentado en el plano IPM (en blanco), y líneas que describen la forma trapezoidal que la circunscribe.

Puesto que el principal interés de esta clase es el de generar regiones potenciales de contener vehículos, estimaremos las diferencias entre regiones para la imagen segmentada y la imagen de referencia en base a las rectas descritas anteriormente. Así la distancia entre el comienzo de un objeto en la imagen segmentada y su homónimo en la imagen de referencia vendrá determinada por la diferencia en píxeles de las rectas horizontales que modelan el comienzo del vehículo.

Por otro lado, para estimar la bondad con la que se detecta el objeto conforme se aleja del origen se obtendrán los ángulos de las rectas que modelan los laterales, así si las diferencias son bajas respecto a la imagen de referencia querrá decir que las regiones son similares tanto en forma como en tamaño.

4.2.3 Clase Línea de carretera

Una línea de carretera se compone de varios segmentos separados los cuales conforman una recta o una curva, y donde cada par de rectas o curvas conformarán un carril. Por tanto, uno de los puntos clave para poder realizar una correcta estimación de carril es detectar estos segmentos correctamente.

Para estimar cómo de bien contribuyen las líneas de carretera detectadas por el sistema, frente a las ideales de la segmentación de referencia se evaluarán dos parámetros:

- ❖ El número de correspondencias correctas entre segmentos de líneas de carretera de la imagen de referencia y la salida del sistema.
- ❖ Diferencias de orientación entre los segmentos de líneas detectados por el sistema y las existentes en la segmentación de referencia.

4.2.4 Secuencias de análisis

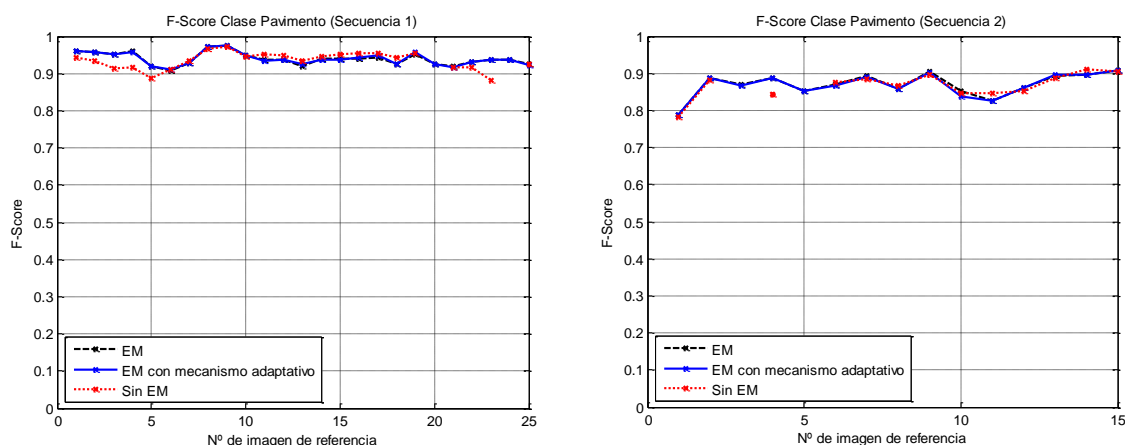
Se va a realizar el análisis sobre dos secuencias de vídeo diferentes. La secuencia 1, de la cual se han etiquetado manualmente 25 imágenes de referencia, presenta un escenario donde los niveles de intensidad se mantienen prácticamente constantes durante todo el vídeo. Por otro lado la secuencia 2, de la que se han generado 15 imágenes de referencia, ha sido incluida para evaluar el rendimiento en condiciones de iluminación variable, ya que dicha secuencia se oscurece a partir de aproximadamente la mitad del vídeo, comprimiendo los valores en la zona inferior del histograma.

4.3 Resultados

A continuación se analizan los obtenidos en la segmentación de las diferentes clases consideradas empleando las dos secuencias indicadas anteriormente.

4.3.1 Clase pavimento

La Figura 4.3 muestra las puntuaciones *F-Score* de la clase pavimento para las dos secuencias, donde se comparan los tres sistemas anteriormente descritos. Como se puede observar, el sistema básico sin EM el cual carece de un mecanismo de control pierde totalmente la detección en las imágenes 19 a 21 y 24 de la secuencia 1 y 2,3 y 5 de la secuencia 2. Esto se debe a dos razones: la detección de pavimento es una etapa crítica en el sistema ya que a partir de ella se modela la probabilidad de pertenecer a cada clase en función de la intensidad. Si en la etapa de estimación no se eliminan correctamente zonas de píxeles pertenecientes a otras clases, la media de la distribución que modela la probabilidad de pertenecer al pavimento se desplaza hacia valores que no son representativos de este, produciendo que un gran número de píxeles de pavimento se clasifiquen como desconocido.



a) *F-Score clase pavimento para la secuencia 1*

b) *F-Score clase pavimento para la secuencia*

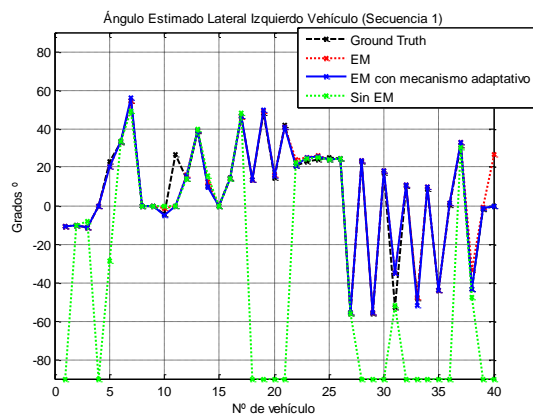
Figura 4.3 Puntuaciones F-Score clase pavimento

Los sistemas con EM son capaces de mantener la detección en estas situaciones gracias a EM y al mecanismo de control. Por un lado, el algoritmo EM presenta cierta robustez frente a malas estimaciones iniciales de los parámetros que modelan las clases. Puesto que la etapa más sensible del sistema es la de estimación de pavimento, ya que la obtención de parámetros del resto de clases dependen de ella, si en esta estimación se incluye una cantidad significativa de elementos pertenecientes a objetos o línea, es muy posible que el sistema clasifique objetos y líneas como pavimento o asigne todos los elementos a la clase desconocido. Es aquí donde destaca la capacidad correctiva de EM, el cual “garantiza que la log-verosimilitud de los datos buenos (con los datos malos marginalizados) crecerá de forma monótona” [21], así aunque se incluyan datos no pertenecientes a la clase pavimento, EM será capaz de corregirlo en cierto grado. Además, en el caso de que la estimación se vea totalmente comprometida y EM no sea capaz de corregirlo, estos sistemas cuentan con el mecanismo de control que detecta dichos fallos, entrando en un estado transitorio donde se usan los últimos parámetros válidos.

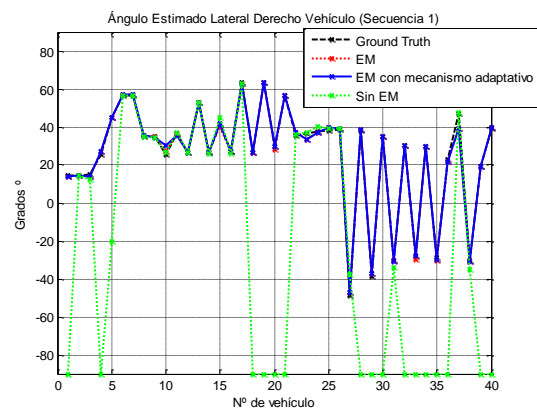
En las imágenes donde la detección es correcta para todos los sistemas no se observa una tendencia que refleje un mejor rendimiento de un sistema respecto a otro. Sin embargo, la secuencia 2 tiene una puntuación media menor causada por el bajo rango dinámico de la secuencia, cuyos valores tienden a estar en la parte izquierda del histograma haciendo que el sistema clasifique las zonas inferiores de pavimento, las cuales son más oscuras, como objeto.

4.3.2 Clase objeto

En las Figuras 4.4 y 4.5 se muestran los ángulos que parametrizan la recta perpendicular al lateral del vehículo. Todas aquellas detecciones a -90° implican que no se han encontrado vehículos en esa imagen.

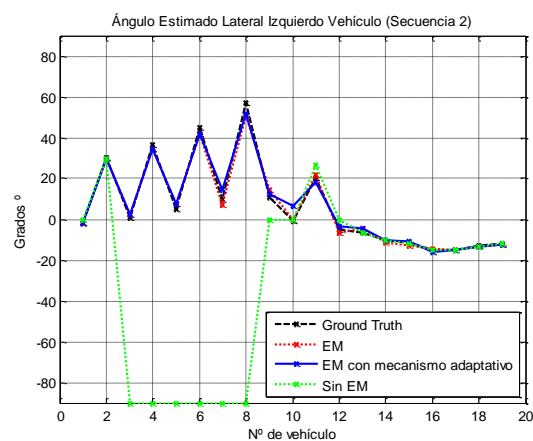


a) Ángulo de la recta que parametriza el lateral izquierdo de los vehículos detectados en la secuencia 1

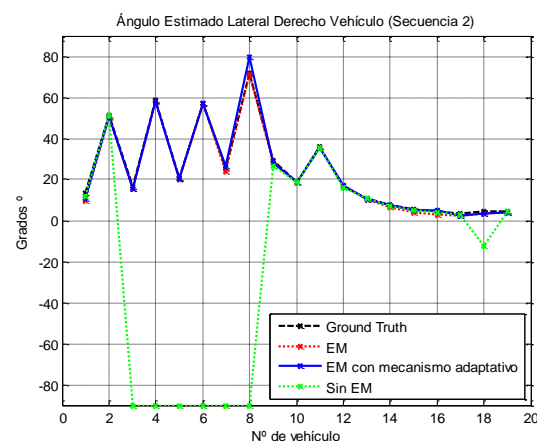


b) Ángulo de la recta que parametriza el lateral derecho de los vehículos detectados en la secuencia 1

Figura 4.4 Ángulos que parametrizan los laterales de los vehículos detectados en la secuencia 1



a) Ángulo de la recta que parametriza el lateral izquierdo de los vehículos detectados en la secuencia 2



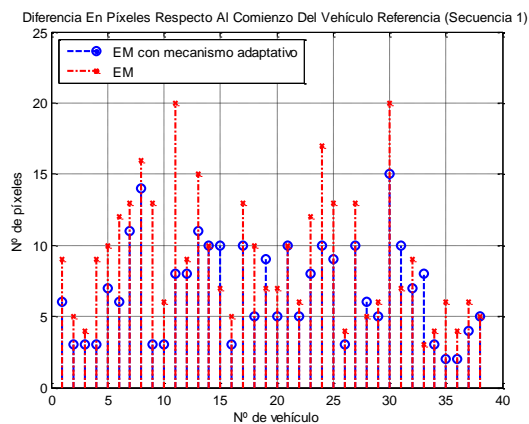
b) Ángulo de la recta que parametriza el lateral derecho de los vehículos detectados en la secuencia 2

Figura 4.5 Ángulos que parametrizan los laterales de los vehículos detectados en la secuencia 2

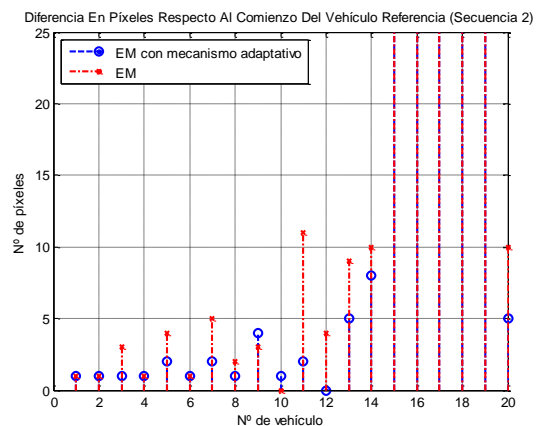
De nuevo, como se puede observar en ambas figuras, el sistema sin EM ni mecanismo de control falla a raíz de las razones expuestas en el apartado anterior, si se realiza una mala estimación del pavimento el resto de clases también se ven comprometidas.

Para la Figura 4.5, a excepción de las de las detecciones 11, 31 (a)) donde existen diferencias respecto de la imagen referencia para ambos sistemas y la detección 40 (a)) en la cual el sistema con EM sin mecanismo no detecta correctamente el lateral del vehículo, los ángulos son iguales o muy similares que las obtenidos en la imagen de referencia. Por otro lado, la Figura 4.5 refleja un mejor funcionamiento del sistema EM sin el control adaptativo, ya que para ambos laterales realiza una detección más precisa, como ilustran las detecciones 7 y 10 de la Figura 4.5 a) y 8 de la Figura 4.5 b).

Respecto a la detección del comienzo del vehículo, se han analizado las diferencias entre el sistema con EM y el sistema EM con mecanismo adaptativo y no se ha incluido el sistema base debido a los numerosos errores que presenta. Las diferencias se ilustran en la Figura 4.6, se observa como el sistema con mecanismo adaptativo tiene una diferencia menor respecto al vehículo de referencia en la mayoría de casos, lo que permite hacer una estimación más precisa de los límites del vehículo, etapa muy importante para la posterior verificación. El gran aumento de las diferencias en las detecciones 14 a 19 de la Figura 4.6 b) son causadas por las condiciones de baja iluminación de esas imágenes, donde la zona inferior presenta unos niveles de intensidad muy similares a los de la sombra vehículo produciendo que en ambos sistemas se clasifiquen esos píxeles como objeto, solapándose el pavimento con la sombra del vehículo delantero.



a) Diferencia en píxeles entre el comienzo de los vehículos detectados y los de referencia para la secuencia 1



b) Diferencia en píxeles entre el comienzo de los vehículos detectados y los de referencia para la secuencia 2

Figura 4.6 Diferencias en píxeles entre el comienzo de los vehículos detectados y los de referencia

El hecho de que para la secuencia 2 el sistema con EM sin mecanismo adaptativo detecte mejor los laterales y no el comienzo del vehículo está muy ligado. La inclusión del mecanismo adaptativo controla la velocidad con la que el sistema se adapta a los cambios de iluminación, lo que causa que en la secuencia 2 la desviación típica de la clase vehículo no varíe tan rápido respecto al sistema sin mecanismo. Esto produce que píxeles con valores de intensidad intermedios entre sombra y pavimento sean clasificados como pavimento en el sistema con mecanismo y como vehículo en el sistema sin mecanismo, en contrapartida, los píxeles del comienzo de los laterales del vehículo se clasifican como desconocido en el sistema con mecanismo al tener una distribución de clase objeto más estrecha, que produce las diferencias entre ángulos estimados para los sistemas.

En resumen, los sistemas con EM y EM con mecanismo adaptativo ofrecen un buen rendimiento en la detección de los laterales del vehículo, siendo además capaces de localizarlo en todas las imágenes, lo que pone de manifiesto la robustez del algoritmo EM unido al mecanismo de control y su utilidad para posteriores aplicaciones de seguimiento. El mecanismo adaptativo considera los parámetros anteriores para la clasificación, filtrando de forma que si alguno se aleja demasiado de la media su influencia se reduce. Esto limita la capacidad de adaptarse a distribuciones puntuales como un cambio de intensidad entre imágenes de las líneas de carretera, aunque añade robustez ante posibles detecciones erróneas no filtradas por el primer nivel.

Además, la detección del comienzo del vehículo se puede refinar limitando la velocidad de adaptación del sistema, aunque es necesario mantener un compromiso para que el resto de elementos no se vean afectados.

4.3.3 Clase línea

Para esta clase se han analizado las líneas del carril por las que circula el vehículo portador de la cámara, se ha desestimado el análisis de las líneas de los carriles adyacentes al presentar altos niveles de ruido que hacían que la distinción de las líneas fuera difícil incluso al generar las imágenes de referencia a mano.

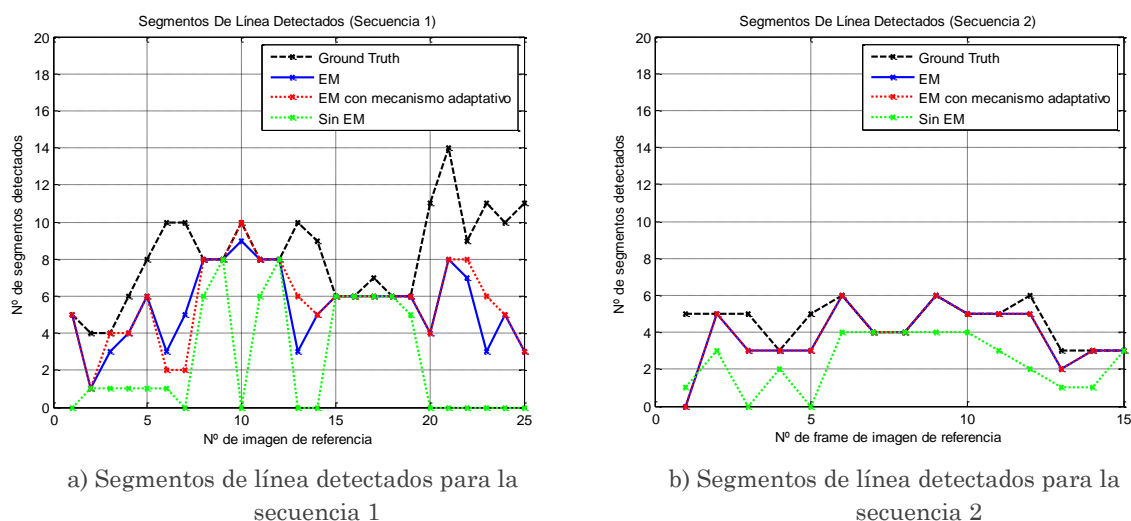
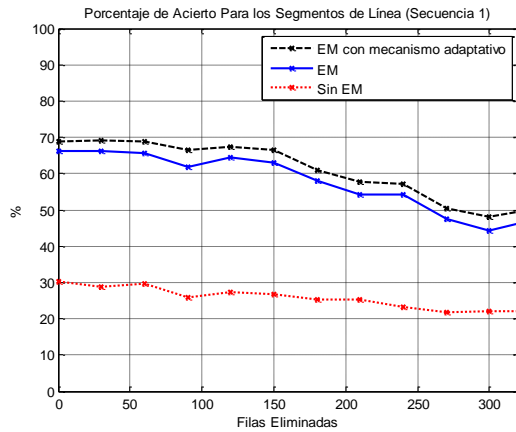


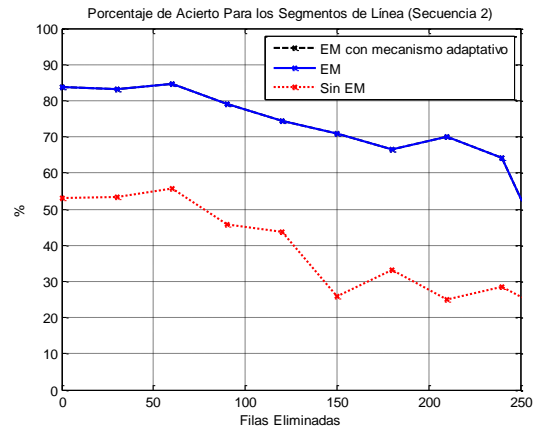
Figura 4.7 Segmentos de línea detectados en los tres sistemas y la imagen de referencia

Este análisis comienza con la comparación del número de segmentos de línea detectados en la imagen de referencia y los diferentes sistemas, ilustrado mediante la Figura 4.7. Para la secuencia 1 de nuevo el sistema más básico es el que peor resultados obtiene, ligado a que en numerosas imágenes pierde la detección, por otro lado el sistema con EM y mecanismo adaptativo obtiene mejores resultados que el sistema EM a excepción de las imágenes 4 y 5. Por otro lado los resultados obtenidos para la secuencia 2 son mejores en general, donde ambos sistemas con EM ofrecen el mismo rendimiento.

Los fallos en las detecciones vienen influenciados en gran parte por las líneas pertenecientes a las zonas más alejadas de la imagen, donde el ruido de la transformación IPM es muy alto. Para ilustrar esto, las Figuras 4.8 y 4.9 muestran respectivamente el porcentaje de acierto y desviación media en grados para toda la secuencia según se eliminan filas de la parte inferior, de forma que las líneas superiores cada vez tienen más influencia.

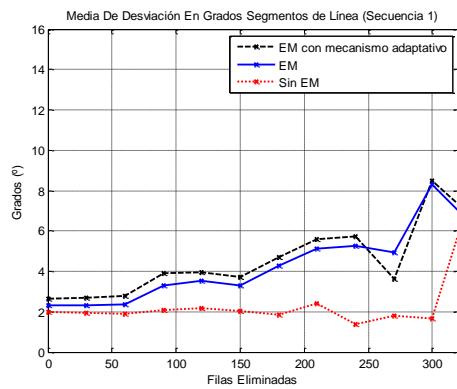


a) Porcentaje de acierto medio para la secuencia 1

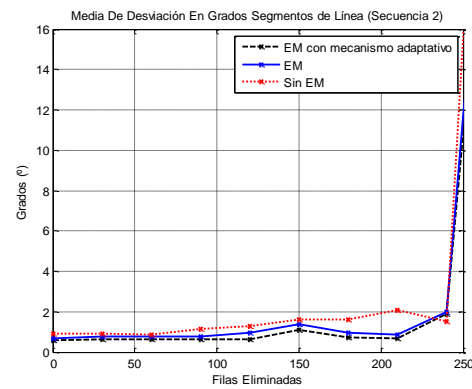


b) Porcentaje de acierto medio para la secuencia 2

Figura 4.8 Porcentaje de acierto para la detección de líneas en función de las filas eliminadas



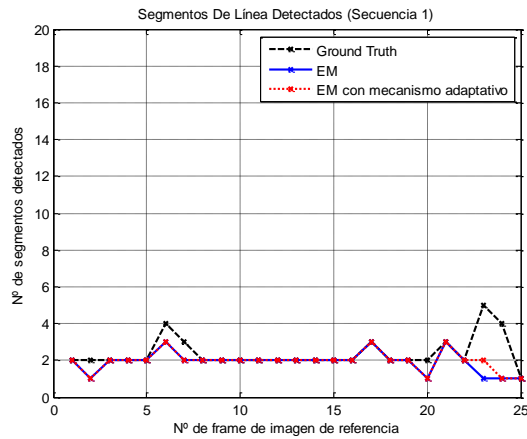
c) Desviación medio para la secuencia 1



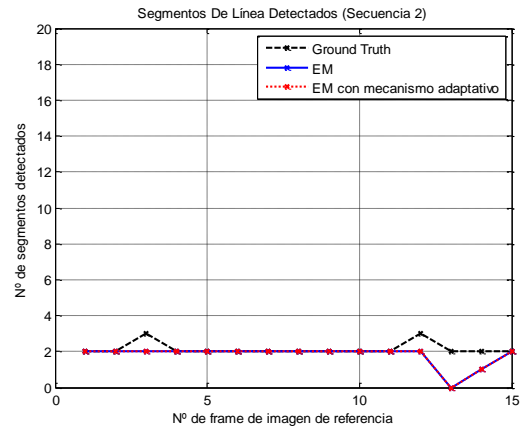
d) Desviación media para la secuencia 2

Figura 4.9 Desviación media en grados de las líneas detectadas respecto de la referencia en función de las filas eliminadas

En ellas puede observarse como disminuye el porcentaje de acierto y se incrementa la media de desviación según eliminamos las líneas más cercanas a la cámara. Cabe destacar que para la secuencia 1 el sistema con mecanismo adaptativo mantiene una media de detección mayor a lo largo de toda la secuencia, mientras que presenta una desviación media mayor. En la secuencia 2 los dos sistemas con EM mantienen el mismo porcentaje de aciertos, mientras que aquel con mecanismo adaptativo presenta una desviación media ligeramente menor.



a) Segmentos de la secuencia 1



b) Segmentos de la secuencia 2

Figura 4. 10 Segmentos detectados para las 50 primeras filas de la imagen

Como conclusión, la detección de las líneas de carretera está muy influenciada por la posición de éstas en la imagen transformada. Para un posterior modelado de carriles se recomienda el uso de las líneas detectadas en la parte inferior de la imagen ya que, como muestra la Figura 4.10 donde se muestran los segmentos para las 50 primeras filas de la imagen, se clasifican con un porcentaje de acierto muy alto, además de que presentan una desviación menor respecto a la imagen de referencia. Esto permite en una etapa posterior realizar un modelado de los carriles más preciso y robusto al detectarse la gran mayoría de líneas en ese entorno y contribuir cada una de ellas con una desviación baja respecto a la referencia contribuir cada línea con una desviación respecto a la referencia muy baja.

4.3.4 Análisis subjetivo

En esta sección se muestran algunas capturas de la salida del sistema con EM y mecanismo adaptativo con el fin de evaluar subjetivamente los resultados obtenidos.

La Figura 4.11 muestra cuatro imágenes correspondientes a la secuencia 1, y la Figura 4.12 de la secuencia 2.

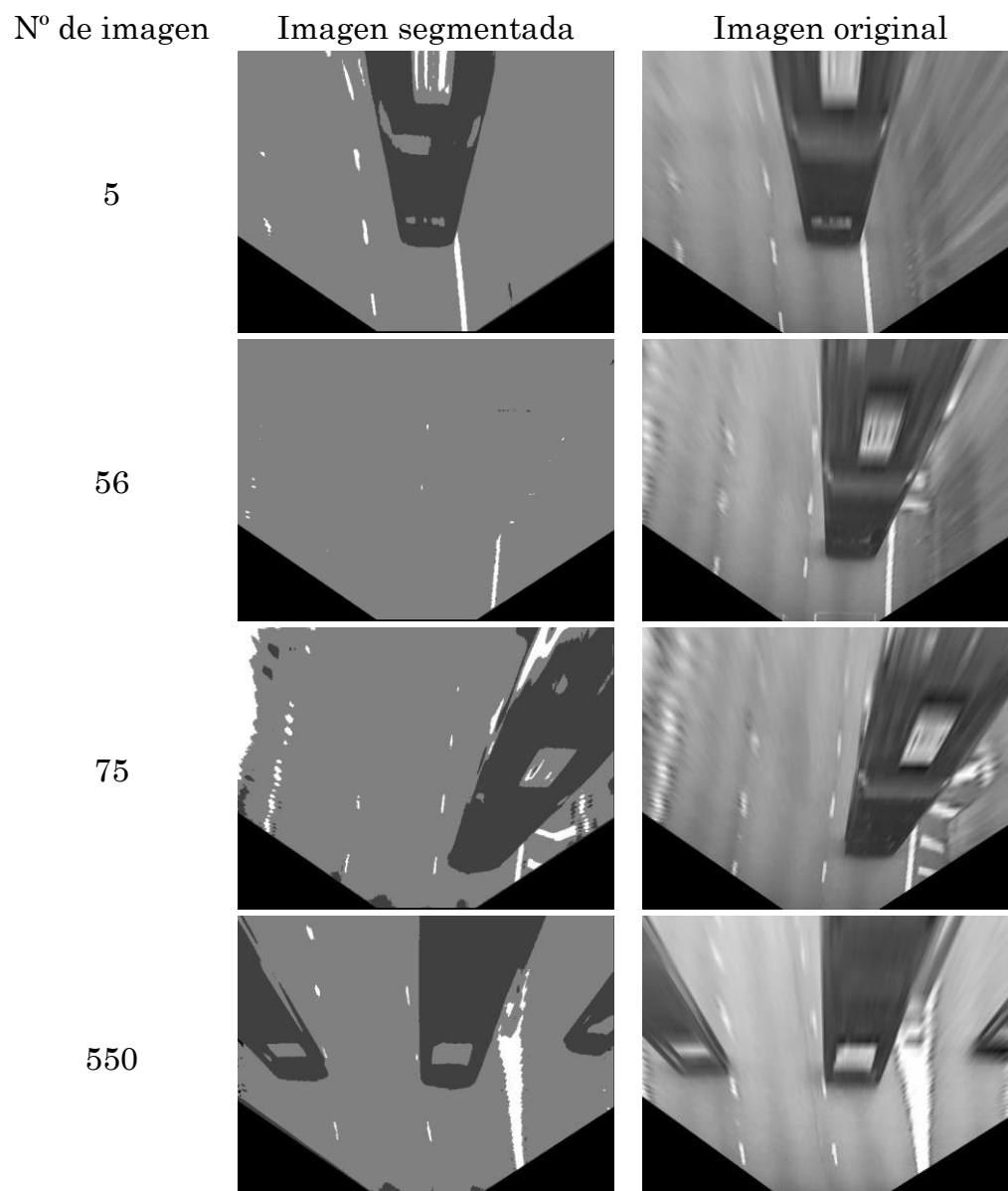


Figura 4. 11 Imágenes de la secuencia 1 segmentadas por el algoritmo

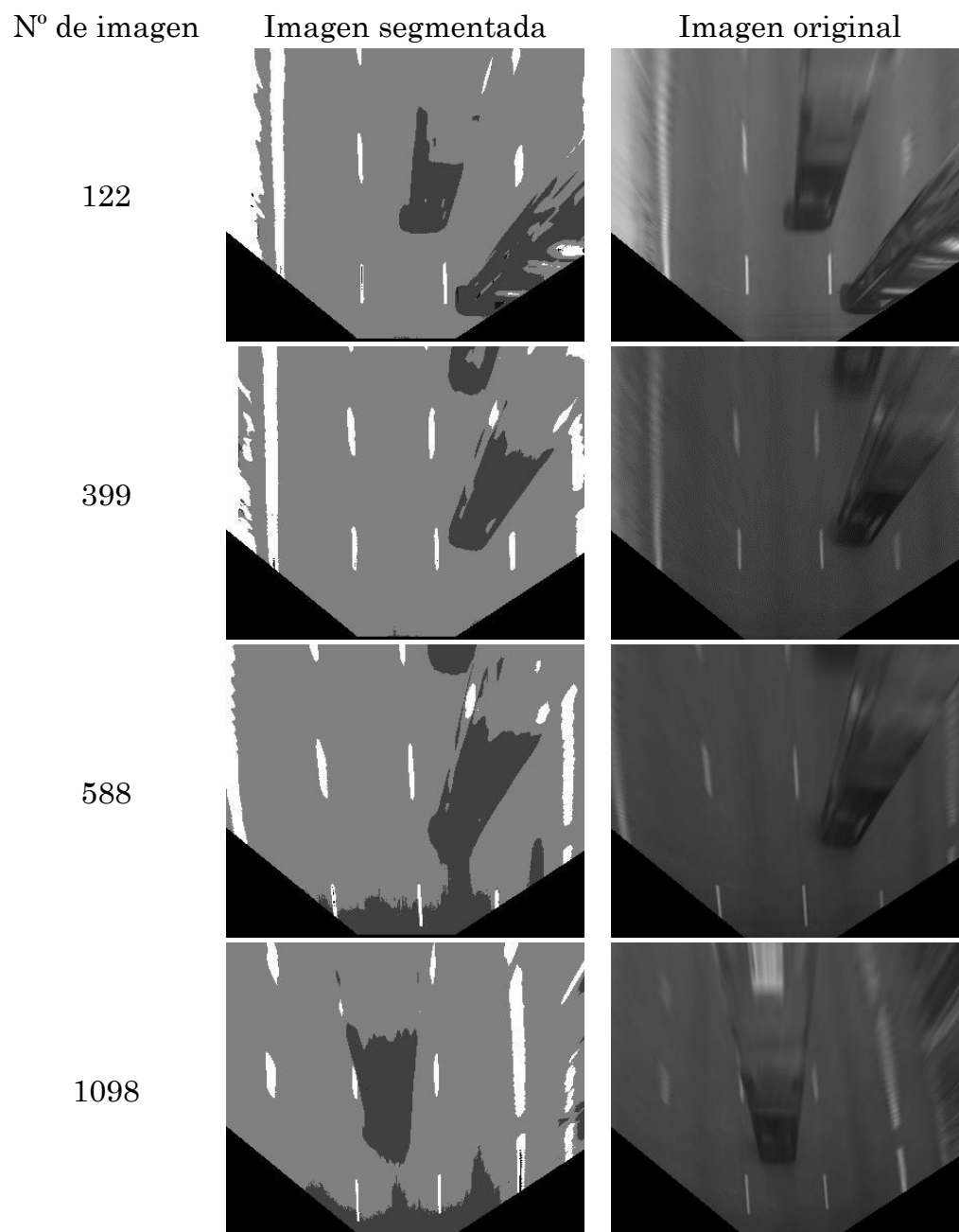


Figura 4. 12 Imágenes de la secuencia 2 segmentadas por el algoritmo

En general, con condiciones de buena luz el sistema realiza una correcta segmentación de los elementos presentes en la imagen. Un visionado de las secuencias muestra que sólo en escasas imágenes se pierde la detección de las sombras, la imagen 56 de la Figura 4.11 muestra un caso donde se pierde la sombra del vehículo y gran cantidad de las líneas.

Cabe destacar que se han clasificado zonas de alta luminancia situadas en los laterales de la imagen que no son realmente línea, estos píxeles deberían haber sido clasificados como desconocido al dar una baja respuesta al filtro de líneas, lo que refleja que el modelado de la clase desconocido no es idóneo para la función de verosimilitud que se obtiene a partir de la respuesta al filtro de líneas, ya que la distribución que modela la respuesta de las líneas presenta una gran desviación, de la misma manera que lo hace la clase desconocido, siendo en ocasiones muy similares. Además es improbable que un elemento que no sea línea ofrezca una respuesta que no sea cercana a cero, por ello se propone en el trabajo futuro un mejor modelado de la clase desconocido para las distribuciones que modelan la respuesta del filtro.

Por otro lado, en condiciones difíciles para el algoritmo, donde la iluminación general de la escena es baja y el pavimento toma unos valores muy oscuros, como muestran las imágenes 588 y 1098 de la Figura 4.12, el sistema es capaz de localizar la sombra del vehículo y detectar todas las líneas de carretera. Presentando el inconveniente de que zonas muy oscuras de pavimento se clasifican como objeto, si bien la parte importante del vehículo se detecta siempre de la misma forma que lo hacen las líneas, lo que pone de manifiesto el correcto funcionamiento en situaciones de iluminación difíciles.

Capítulo 5: Conclusiones y trabajo futuro

El objetivo del sistema desarrollado es segmentar las imágenes provenientes de una cámara situada en el frontal de un vehículo para la detección de líneas de carretera, pavimento y vehículos. El análisis de las pruebas arroja unos buenos resultados para los sistemas que incluyen el algoritmo *Expectation-Maximization* y el mecanismo de control, lo que pone de manifiesto la necesidad de su uso para mejorar y añadir robustez al sistema. Además se han mostrado los efectos del ruido de la transformación IPM en las zonas más alejadas de la imagen, donde la información disponible deja de ser fiable.

Además, por la forma de clasificación que se ha usado para la clase línea, donde se utilizan filtros morfológicos específicamente diseñados para ello se puede asumir que las zonas detectadas son realmente líneas de carretera. Los vehículos, sin embargo, son clasificados únicamente en base a su sombra, por lo que dichas regiones, aunque tengan una alta probabilidad de contener vehículos, son hipótesis que han de ser verificadas. Por estas razones se plantean los siguientes puntos de trabajo futuro:

Mejoras sobre el sistema:

- ❖ Introducir, si esta estuviera disponible, información sobre la dirección del vehículo en el filtrado direccional para la generación de probabilidades *a priori*.
- ❖ Mejorar la pertenencia a la clase desconocido para las funciones de verosimilitud obtenidas a partir de la respuesta al filtro de líneas.
- ❖ Modelar la incertidumbre introducida por el ruido de la transformación IPM mediante cambios en la probabilidad *a priori* en función de la zona de la imagen.

- ❖ Realizar una estimación adaptativa del ancho de línea para actualizar el parámetro del filtro de líneas de forma dinámica.
- ❖ Realizar una estimación adaptativa del parámetro α del mecanismo adaptativo en función de la variabilidad de las imágenes.
- ❖ No realizar una estimación de las distribuciones en todas las imágenes para aliviar el coste computacional.

Implementaciones a partir del trabajo realizado:

- ❖ Implementación del sistema completo que integre la transformación IPM y el sistema descrito para tener información disponible en los dos dominios.
- ❖ Comprobación de las hipótesis de vehículos generadas por el sistema en el dominio original mediante reconocimiento de patrones y aprendizaje máquina.
- ❖ Modelado del carril a partir de la segmentación de las líneas de carretera obtenida.

Referencias

- 1.Sun, Z., Bebis, G., & Miller, R. (2006). On-road vehicle detection: A review.*Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(5), 694-711.
- 2.Alvarez, J. M., LeCun, Y., Gevers, T., & Lopez, A. M. (2012, January). Semantic road segmentation via multi-scale ensembles of learned features. In *Computer Vision–ECCV 2012. Workshops and Demonstrations* (pp. 586-595). Springer Berlin Heidelberg.
- 3.Nieto, M., Laborda, J. A., & Salgado, L. (2011). Road environment modeling using robust perspective analysis and recursive Bayesian segmentation.*Machine Vision and Applications*, 22(6), 927-945.
- 4.Kluge, K., & Lakshmanan, S. (1995, September). A deformable-template approach to lane detection. In *Intelligent Vehicles' 95 Symposium., Proceedings of the* (pp. 54-59). IEEE.
- 5.Wang, Y., Teoh, E. K., & Shen, D. (2004). Lane detection and tracking using B-Snake. *Image and Vision computing*, 22(4), 269-280.
- 6.Wang, Y., Shen, D., & Teoh, E. K. (1998, October). Lane detection using catmull-rom spline. In *IEEE International Conference on Intelligent Vehicles*(pp. 51-57).
- 7.Parajuli, A., Celenk, M., & Riley, H. B. (2013). Robust lane detection in shadows and low illumination conditions using local gradient features. *Open Journal of Applied Sciences*, 3(01), 68.
- 8.Gonzalez, J. P., & Özgüner, Ü. (2000). Lane detection using histogram-based segmentation and decision trees. In *Intelligent Transportation Systems, 2000. Proceedings. 2000 IEEE* (pp. 346-351). IEEE.
- 9.Bertozzi, M., & Broggi, A. (1998). GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection. *Image Processing, IEEE Transactions on*, 7(1), 62-81.
- 10.Arróspide, J., Salgado, L., & Nieto, M. (2013, August) A Review of Computer Vision Techniques for On-Road Vehicle Detection and Tracking. Preprint submitted to Expert Systems with Applications

11. Kormann, B., Neve, A., Klinker, G., & Stechele, W. (2010). Stereo Vision based Vehicle Detection. In *VISAPP (2)* (pp. 431-438).
12. Mori, H., & Charkari, N. M. (1993). Shadow and rhythm as sign patterns of obstacle detection. In *Industrial Electronics, 1993. Conference Proceedings, ISIE'93-Budapest., IEEE International Symposium on* (pp. 271-277). IEEE.
13. Bertozzi, M., Broggi, A., & Fascioli, A. (1998). Stereo inverse perspective mapping: theory and applications. *Image and vision computing*, 16(8), 585-590.
14. Nieto, M., Salgado, L., Jaureguizar, F., & Cabrera, J. (2007, June). In *Intelligent Vehicles Symposium, 2007 IEEE* (pp. 315-320). IEEE.
15. Bertozzi, M., Broggi, A., Fascioli, A., & Nichele, S. (2000, October). Stereo vision-based vehicle detection. In *IEEE Intelligent Vehicles Symposium* (pp. 39-44).
16. Srinivasa, N. (2002, June). Vision-based vehicle detection and tracking method for forward collision warning in automobiles. In *Intelligent Vehicle Symposium, 2002. IEEE* (Vol. 2, pp. 626-631). IEEE.
17. Lee, D., & Park, Y. (2006, August). Robust vehicle detection based on shadow classification. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* (Vol. 3, pp. 1167-1170). IEEE.
18. Matthews, N. D., An, P. E., Charnley, D., & Harris, C. J. (1996). Vehicle detection and recognition in greyscale imagery. *Control Engineering Practice*, 4(4), 473-479.
19. Baş, E., & Salman, F. S. (2007, June). Automatic vehicle counting from video for traffic flow analysis. In *Intelligent Vehicles Symposium, 2007 IEEE* (pp. 392-397). Ieee.
20. Naito, T., Ito, T., & Kaneda, Y. (2007, June). The obstacle detection method using optical flow estimation at the edge image. In *Intelligent Vehicles Symposium, 2007 IEEE* (pp. 817-822). IEEE.
21. Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.
22. Gonzalez, R. C. (2009). *Digital image processing*. Pearson Education.